

1 **ABSTRACT**

2 A linguistic disambiguation system and method creates a knowledge base
3 by training on patterns in strings that contain ambiguity sites. The string patterns
4 are described by a set of reduced regular expressions (RREs) or very reduced
5 regular expressions (VRREs). The knowledge base utilizes the RREs or VRREs
6 to resolve ambiguity based upon the strings in which the ambiguity occurs. The
7 system is trained on a training set, such as a properly labeled corpus. Once
8 trained, the system may then apply the knowledge base to raw input strings that
9 contain ambiguity sites. The system uses the RRE- and VRRE-based knowledge
10 base to disambiguate the sites.